Reliability of American Society of Anesthesiologists physical status classification

INTRODUCTION

The American Society of Anesthesiologists Physical Status (ASA-PS) classification is a widely used grading system for the pre-operative health of a surgical patient. It was originally developed in 1941 by Saklad *et al*. and then modified in 1961 by Dripps *et al*.[1] into a five-class version.

According to other researchers, the ASA-PS classification should be modified and adapted to the paediatric population because there are many differences between the physiology and pathology of adults and children. Many studies have tested the relation between the ASA classification and several outcomes[2,3] such as mortality, cardiac arrest, morbidity, length of stay and predictors of blood loss. The reliability of the ASA-PS classification has been widely evaluated,[4,5] but there are different conclusions on the ASA-PS classification reliability. There is no agreement on the level of reliability of the scale.

We conducted a systematic review on the state of studies on the reliability of the ASA-PS classification. To our knowledge, there is only one review on the ASA-PS classification,[6] and there are no systematic reviews on its reliability.

The primary aim was to check the state of studies on the reliability of the ASA-PS classification for the broad population of adults and children waiting for surgery.

METHODS

The questions for the review were as follows: (1) What is the level of reliability of the ASA-PS system among the selected studies? (2) How is the quality of reporting among published studies on reliability of the ASA-PS system? (3) How is the quality of statistical methodology among published studies on reliability of the ASA-PS system?

We used the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Guideline for the first part of the review protocol, the selection of studies. Then, we used a modified version of the Standards for the Reporting of Diagnostic Accuracy (STARD)[7] and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Guidelines (http://www.strobe-statement.org) to analyse the quality of reporting among studies selected [Appendix 1 and 2]. Finally, we used the Statistical Analyses and Methods in the Published Literature (SAMPL) Guidelines (http://www.equator-network.org/reporting-guidelines/sampl) to test the quality of statistical methodology among the collected studies.

The outcome measures were reliability tested using the k statistic, Cohen's kappa or the intra-class correlation coefficient; the percentage of STARD, STROBE, SAMPL items respected.

The inclusion and exclusion criteria and the methodology to assess the risk bias (of individual studies and of the cumulative evidence of included studies) are shown in Appendix 3.

The systematic search of the international literature published from 1941 through 30 November 2014 was performed using keywords and strategy as shown in Appendix 4. After literature searches, we found 693 records. The selection of articles included in the review was performed in a three-phase process [Figure 1] according to PRISMA Guidelines and is shown in the Appendix 3.

RESULTS

We collected 13 studies for final analysis [Table 1]. Three of the studies collected were conducted on a paediatric population and only two were done on real patients. The prevalent study design was observational with scenarios. We found only one multi-centric study.

The ASA-PS classification reliability was tested among anaesthesiologists and nurses from public and private hospitals.

Eight of the 13 studies tested the inter-rater reliability using the kappa statistic, but the researchers used a very heterogeneous statistical methodology namely, un-weighted, weighted, quadratic kappa. According to the Landis and Koch terminology and classification of kappa value,[8] two of the eight studies found fair inter-rater reliability (k range = 0.21–0.40), three had moderate reliability (k range = 0.47–0.53) and three had good reliability (k range = 0.61–0.82).

All studies conducted on children found a moderate (k range = 0.47–0.50) inter-rater reliability.

Seven studies respected less than 60% of STARD items, and eight respected less than 60% of STROBE and SAMPLE items [Table 1].

None of the studies selected met all 25 items of the STARD, STROBE and SAMPL guidelines.

The studies that met more items of the STARD and STROBE checklists were those of Ringdal *et al.*, Sankar *et al.* and Cuvillon *et al.*[5,9,10] All of these studies respected more than 80% of items of both checklists [Table 1]; however, only Sankar *et al.* and Cuvillon *et al.* respected more than 80% of SAMPL items.

DISCUSSION

In this review, the ASA-PS classification shows a wide inter-rater agreement range among all studies included,

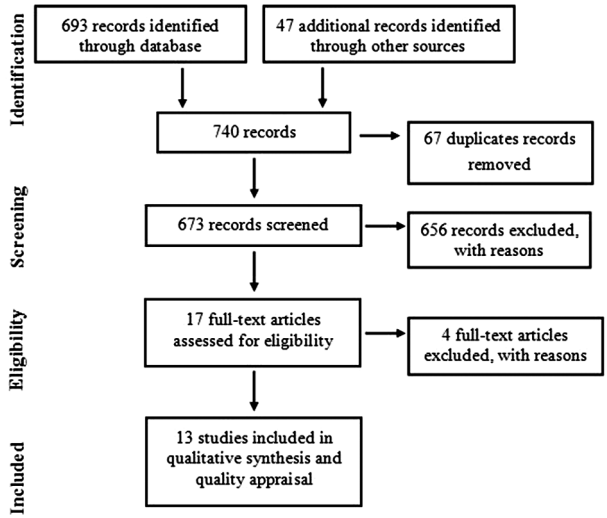

Figure 1: Review process

Table 1: Characteristics of studies selected

Author	Study Design	Target Population	Outcome	Results	*Modified STARD items Met percentage (*n*)	°Modified STROBE items met percentage (*n*)	§Modified SAMPL items met percentage (*n*)
Owens W *et al*. 1978. USA	Observational	10 scenarios: Adult 199 raters	Consistency		53 (10)	43 (13)	40 (16)
Haynes SR *et al*. 1995. USA	Observational	10 scenarios: Adult 97 raters	Consistency		42 (8)	40 (12)	38 (15)
Ranta S *et al*. 1997. Finland	Observational	10 scenarios: Adult patients. 108 raters	Variation in ASA Class allocation		53 (10)	43 (13)	40 (16)
Mak PH *et al*. 2002. Hong Kong	Observational	10 scenarios: Adult 97 raters	Inter-rater reliability	Fair inter-rater reliability	63 (12)	40 (12)	38 (15)
Aronson CWL *et al*. 2003. USA	Observational	10 scenarios: Adult patients 70 raters	Consistency		58 (11)	40 (12)	38 (15)
Ragheb J *et al*. 2006. USA	Observational	10 scenarios: Pediatric 54 raters	Inter-rater reliability	Moderate agreement Agreement excellent for all ASA classes	47 (89)	47 (14)	45 (18)
Aplin S *et al*. 2007. Australia	Observational	15 scenarios: Pediatric 130 rates	Inter-rater reliability	Poor inter-rater reliability	58 (11)	57 (17)	55 (22)
Burgoyne LL *et al*. 2007. USA	Observational	10 scenarios: Pediatric 267 raters	Inter-rater reliability	Moderate agreement	84 (16)	73 (22)	70 (28)
Cuvillon *P et al*. 2011. France, Canada	Multicenter Cross-over Observational	1,554 real patients 2 raters	Inter-rater reliability	Moderate agreement Poor consistency of ASA application	84 (16)	93 (28)	88 (35)
Ringdal KG *et al*. 2013. Norway	Observational	50 scenarios: Adult 19 raters	Inter-rater reliability	Inter-rater reliability from moderate to substantial	90 (17)	80 (24)	75 (30)
Iherjirika RC *et al*. 2014.	Observational	9 scenarios: Adult 33 raters	Inter-rater reliability	ASA-PS reliability was substantial	79 (15)	63 (19)	60 (24)
Sankar A *et al*. 2014. Canada	Retrospective	10,864 real patients: Adults. 2 raters	Inter-rater reliability Validity pred. mort.	ASA-PS has a moderate inter-rater reliability	90 (17)	87 (26)	85 (34)
Riley R *et al*. 2014.	Observational	10 scenarios: Adult 151 raters	Inter-rater reliability	ASA-PS has a fair inter-rater reliability	47 (89)	50 (15)	45 (18)

*STARD – The original version of STARD included 25 Items; the modified STARD version (showed in the table) includes 19 items; °STROBE – The original version of STROBE included 22 Items and 12 "sub-items" (total 34); the modified STROBE (showed in the table) version includes 22 items and 8 "sub-items" (total 30); §SAMPL – The modified SAMPL version (showed in the table) includes 40 items

from fair to very good agreement; however, there was a prevalence of moderate agreement. Seven of the nine studies reported a k inter value higher than 0.4.

Because there are limited data on intra-rater reliability for ASA-PS classification, we think future studies should be planned on these topics. Finally, there are few data on the reliability for patients included in ASA Classes V and VI and limited data on the ASA-PS scale performance with younger children, but it shows moderate agreement with the research available.

We chose to plan a review on the ASA-PS classification reliability because the reliability is a fundamental characteristic of clinical scale.

There is an inter-rater reliability and intra-rater reliability for clinical scores. They are usually analysed using the k statistic, Cohen's kappa.

The wide inter-rater agreement range among all studies included could be explained by the fact that there is a wide discrepancy on the statistical methodology used.

Among the studies collected, we found a prevalence of moderate agreement for the ASA-PS classification; this does not mean that the classification has bad performance, but it could be caused by the bad educational training of the raters.

Furthermore, in our opinion, many previous studies on ASA-PS reliability have several limitations in the methodology: very few studies used a statistical methodology to estimate the right sample size of scenarios or patients; many studies used very few scenarios; finally, almost all of the previous studies used paper scenarios instead of real patients.

The quality of future research on ASA-PS classification should improve: there need to be prospective multi-centre studies based on real patients, planned with a better statistical methodology.

CONCLUSION

The ASA-PS classification seems to have a wide range of inter-rater agreement with a prevalence of moderate value. The administrative staff should be careful to use the ASA-PS classification for administrative billing procedures because of its heterogeneous reliability. Moreover, the physicians should consider the moderate and wide range of agreement of the classification when they use it for general communications. Ideally before using the ASA-PS classification, a test on its reliability among the users should be performed.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

Nicola Parenti, Maria Letizia Bacchi Reggiani[1], Daniela Percudani[2], Rita Maria Melotti[3]

Department of Anesthesia, Hospital of Imola, [1]Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, [2]Librarian, University of Parma, [3]Department of Medical and Surgical Sciences, University of Bologna, Italy

Address for correspondence:
Dr. Nicola Parenti,
Via San Vitale 96, 40125 Bologna, Italy.
E-mail: nipar71@yahoo.it

REFERENCES

1. Dripps RD, Lamont A, Eckenhoff JE. The role of anesthesia in surgical mortality. JAMA 1961;178:261-6.
2. Tiret L, Hatton F, Desmonts JM, Vourc'h G. Prediction of outcome of anaesthesia in patients over 40 years: A multifactorial risk index. Stat Med 1988;7:947-54.
3. Wolters U, Wolf T, Stützer H, Schröder T. ASA classification and perioperative variables as predictors of postoperative outcome. Br J Anaesth 1996;77:217-22.
4. Haynes SR, Lawler PG. An assessment of the consistency of ASA physical status classification allocation. Anaesthesia 1995;50:195-9.
5. Sankar A, Johnson SR, Beattie WS, Tait G, Wijeysundera DN. Reliability of the American Society of Anesthesiologists physical status scale in clinical practice. Br J Anaesth 2014;113:424-32.
6. Daabiss M. American Society of Anaesthesiologists physical status classification. Indian J Anaesth 2011;55:111-5.
7. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, *et al.* The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med 2003;138:W1-12.
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
9. Ringdal KG, Skaga NO, Steen PA, Hestnes M, Laake P, Jones JM, *et al.* Classification of comorbidity in trauma: the reliability of pre-injury ASA physical status classification. Injury 2013;44:29-35.
10. Cuvillon P, Nouvellon E, Marret E, Albaladejo P, Fortier LP, Fabbro-Perray P, *et al.* American Society of Anesthesiologists' physical status system: a multicentre Francophone study to analyse reasons for classification disagreement. Eur J Anaesthesiol 2011;28:742-7.

Access this article online	
Quick response code	Website: www.ijaweb.org
	DOI: 10.4103/0019-5049.177875

How to cite this article: Parenti N, Reggiani ML, Percudani D, Melotti RM. Reliability of American Society of Anesthesiologists physical status classification. Indian J Anaesth 2016;60:208-14.

Appendix 1: Modified STARD guidelines

Section and topic	Item#	
Title/abstract/ keywords	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity')
Introduction	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups
Methods		
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where data were collected
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?
Test methods	7	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard
	8	Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard
Statistical methods	9	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals)
	10	Methods for calculating test reproducibility, if done.
Results		
Participants	11	When study was performed, including beginning and end dates of recruitment
	12	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms)
	13	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended)
Test results	14	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.
Estimates	15	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals)
	16	How indeterminate results, missing data and outliers of the index tests were handled.
	17	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done
	18	Estimates of test reproducibility, if done
Discussion	19	Discuss the clinical applicability of the study findings

We deleted the items number 7, 10, 11, 17, 19, and 20 of original STARD version and we developed the STARD modified version

Appendix 2: Modified STROBE checklist

	Item no	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	*Cohort study*-Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up
		Case-control study-Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls
		Cross-sectional study-Give the eligibility criteria, and the sources and methods of selection of participants
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at

Contd...

Appendix 2: Contd...

	Item no	Recommendation
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding
		(b) Describe any methods used to examine subgroups and interactions
		(c) Explain how missing data were addressed
		(d) *Case-control study*-If applicable, explain how matching of cases and controls was addressed
		Cross-sectional study-If applicable, describe analytical methods taking account of sampling strategy
		(e) Describe any sensitivity analyses
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study-e.g. numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (e.g. demographic, clinical, social) and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
Outcome data	15*	*Cohort study*-Report numbers of outcome events or summary measures over time
		Case-control study-Report numbers in each exposure category, or summary measures of exposure
		Cross-sectional study-Report numbers of outcome events or summary measures
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g. 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(b) Report category boundaries were categorized when continuous variables
Other analyses	17	Report other analyses done-e.g. analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies

APPENDIX 3. METHODS

We included all studies on the reliability of ASA conducted on all ages of patients in all languages.

To assess the risk of bias of individual studies included, pairs of reviewers worked independently and determined the adequacy of randomization , blinding of patients, health care providers, data collectors, and outcome assessors.

According to PRISMA guidelines, to assess the risk of bias that may affect the cumulative evidence of included studies we compared the outcomes in the studies' protocol (or listed in methods section) with the published reports in the result section.

We excluded the duplicate studies and the studies that did not include reliability as an outcome measure.

We did not perform any pre-specified analyses for assessing risk of bias across studies because too few included studies.

The three reviewers' yes/no agreement for each study were entered into an Excel 2010 (Microsoft Corporation) spreadsheet, and Fleiss' kappa for observed agreement was performed. We obtained a Fleiss' kappa score of k=0,70, equating to a "substantial" level of agreement between the raters .

Review process [Figure 1].

In the first phase, one author, an expert in literature research, conducted a literature search of the following databases: PubMed, Embase, Cochrane Library, Web of Science and Scopus. In the second phase, three researchers independently, blind to the assignment of each other, performed a screening by regarding

eligibility criteria of the five lists of articles (title and abstract) selected in phase one. All duplicates were removed. In this way, potentially useful articles were selected with their full text. In the third phase, three researchers independently examined all full-text articles to select the studies that met the inclusion criteria for the systematic review.

APPENDIX 4.LITERATURE STRATEGY RESEARCH

PubMed

Search strategy with Mesh 119 records retrieved:

#1 Search "Patients/classification"[Mesh] OR "Health Status"[Mesh] OR "Health Status Indicators" [Mesh:noexp] OR "Health Surveys" [Mesh:noexp] OR "Preoperative Care"[Mesh:noexp] OR "Physical Examination"[Mesh:noexp] OR "Health Status"[Mesh:noexp] AND (("Anesthesiology"[Mesh] OR "Anesthesia"[Mesh:noexp]) AND (asa OR American society of anesthesiologists OR American society of anaesthesiologists))

Search strategy with Text words 158 records retrieved:

#2 Search (physical OR score* OR scoring OR criteria OR grade* OR classification* OR class OR index OR parameter* OR standard* OR grading) AND (american society anesthesiologists OR american society anaesthesiologists OR ASA) Field: Title

Previouses strategies combined with OR 239 records retrieved:

#3 Search #1 OR #2

EMBASE

Search strategy with EmTree 102 records retrieved:

#1 Search ('anesthesia'/de OR 'anesthesist'/de OR 'anesthesiology'/de AND ('patient'/de OR 'health status'/de OR 'health survey'/exp OR 'physical examination'/de OR 'preoperative evaluation'/de OR 'classification'/de OR 'functional status'/de OR 'procedures'/de)) AND (asa OR "American society of anesthesiologists" OR "American society of anaesthesiologists")

Search strategy with Text words 168 records retrieved:

#2 Search (physical OR score* OR scoring OR criteria OR grade* OR classification* OR class OR index OR parameter* OR standard* OR grading AND (american society anesthesiologists OR american society anaesthesiologists OR ASA)):ti Previouses strategies combined with OR 220 records retrieved:

#3 Search #1 OR #2

Web of Science

Search strategy with Text words 178 records retrieved:

#1 Search Field title: (physical OR score* OR scoring OR criteria OR grade* OR classification* OR class OR index OR parameter* OR standard* OR grading) AND (american society anesthesiologists OR ASA OR american society anaesthesiologists) 178 records

Scopus

Search strategy with Text words 56 records retrieved:

#1 Search Field title: (physical OR score* OR scoring OR criteria OR grade* OR classification* OR class OR index OR parameter* OR standard* OR grading) AND (american society anesthesiologists OR ASA OR american society anaesthesiologists)

Cochrane Library

Search strategy with Text words none records retrieved:

#1 Search Field title: american society of anesthesiologists OR ASA OR american society of anaesthesiologists.